

PD - 12/3/98, FD - 9/28/1999

177

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2002, 13:48:36 ; Search time 1341.91 Seconds
(without alignments)
1954.715 Million cell updates/sec

Title: US-09-407-804A-8

Perfect score: 159

Sequence: 1 atgttaaccacaaagaattttt.....ccgctatcgctgaattataa 159

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_on.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_on.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_pat.*
- 23: em_ph.*
- 24: em_pl.*
- 25: em_ro.*
- 26: em_sts.*
- 27: em_sy.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_htgo_hum.*
- 31: em_htgo_inv.*
- 32: em_htgo_rod.*
- 33: em_htg_hum.*
- 34: em_htg_inv.*
- 35: em_htg_rod.*
- 36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Match	Length	DB	ID	Description
1	149.4	94.0	42942	7	AB045978	AB045978 Staphyloc
2	143	89.9	348527	1	AP003360	AP003360 Staphyloc
3	52.6	33.1	43081	7	AP001553	AP001553 Bacterioph
4	40	25.2	45636	7	AB044554	AB044554 Staphyloc
5	36.8	23.1	4620	9	AB037782	AB037782 Homo sapi
6	36.8	23.1	25943	3	CEC386C	Z93375 Caenorhabdi
7	36.8	23.1	161198	2	AC015867	AC015867 Homo sapi
8	36.8	23.1	175271	2	AC090898	AC090898 Homo sapi
9	36.8	23.1	188346	2	AC010189	AC010189 Homo sapi
10	36.8	23.1	189009	2	AC068588	AC068588 Homo sapi
11	35.8	22.5	149109	2	AC023181	AC023181 Homo sapi
12	35.8	22.5	149246	9	AC066580	AC066580 Homo sapi
13	35.8	22.5	207600	9	AC024910	AC024910 Homo sapi
14	35.8	22.5	225542	9	AC022379	AC022379 Homo sapi
15	35.6	22.4	105222	8	AC006220	AC006220 Arabidops
16	35.2	22.1	171200	9	AC090946	AC090946 Homo sapi
17	35.2	22.1	171543	33	AC019245	AC019245 Homo sapi
18	35.2	22.1	179717	9	AC087858	AC087858 Homo sapi
19	34.8	21.9	166065	2	AC013552	AC013552 Homo sapi
20	34.8	21.9	171698	2	AL359175	AL359175 Homo sapi
21	34.8	21.9	173852	2	AL158160	AL158160 Homo sapi
22	34.4	21.6	153988	2	AC091530	AC091530 Papio cyn
23	34.4	21.6	200964	2	AC091658	AC091658 Papio cyn
24	33.6	21.1	87235	2	AC022963	AC022963 Homo sapi
25	33.6	21.1	134841	9	AC016689	AC016689 Homo sapi
26	33.4	21.0	3389	9	AK022494	AK022494 Homo sapi
27	33.4	21.0	3635	9	HUMORFKG10	D31886 Human mRNA
28	33.2	20.9	72628	2	AC092306	AC092306 Homo sapi
29	33.2	20.9	173157	2	AC026457	AC026457 Homo sapi
30	33.2	20.9	173354	2	AC016478	AC016478 Homo sapi
31	33.2	20.9	193804	2	AC092350	AC092350 Homo sapi
32	33	20.8	7082	1	AF001497	AF001497 Campyloba
33	33	20.8	100170	9	AL161743	AL161743 Human DNA
34	33	20.8	133054	2	AL445200	AL445200 Homo sapi
35	33	20.8	138756	2	AL359814	AL359814 Homo sapi
36	33	20.8	158641	9	AL359541	AL359541 Human DNA
37	33	20.8	169779	2	AL450995	AL450995 Homo sapi
38	33	20.8	282183	1	CJ11168X4	AL390777 Campyloba
39	32.8	20.6	162714	2	AC027569	AC027569 Homo sapi
40	32.8	20.6	168637	9	AC073137	AC073137 Homo sapi
41	32.6	20.5	1910	9	BC012222	BC012222 Homo sapi
42	32.6	20.5	130513	9	AC004592	AC004592 Homo sapi
43	32.6	20.5	141756	2	AL358774	AL358774 Mus muscu
44	32.6	20.5	196444	2	AC022854	AC022854 Homo sapi
45	32.4	20.4	152094	2	AL591594	AL591594 Homo sapi

ALIGNMENTS

RESULT	1
LOCUS	AB045978 42942 bp DNA PHG 26-MAY-2001
DEFINITION	Staphylococcus aureus temperate phage phisLT genomic DNA, complete sequence.
ACCESSION	AB045978
VERSION	AB045978.1 GI:12697822
KEYWORDS	
SOURCE	Staphylococcus aureus temperate phage phisLT DNA.
ORGANISM	Staphylococcus aureus temperate phage phisLT
REFERENCE	1 (sites)
AUTHORS	Narita,S., Kaneko,J., Chiba,J., Piemont,Y., Jarraud,S., Etienne,J. and Kamio,Y.
TITLE	Phage conversion of Pantone-Valentine leukocidin in Staphylococcus aureus: molecular analysis of a pVL-converting phage, φSLT
JOURNAL	Gene 268 (1-2), 195-206 (2001)
MEDLINE	21261956
REFERENCE	2 (bases 1 to 42942)
AUTHORS	Kaneko,J., Narita,S. and Kamio,Y.

TITLE Direct Submission
JOURNAL Submitted (12-JUL-2000) Jun Kaneko, Tohoku University, Graduate
school of Agricultural Science; 1-1 Tsutsumidori Amamiyamachi,
Sendai, Miyagi 981-8555, Japan
(E-mail: jkaneko@biochem.tohoku.ac.jp, Tel: 81-22-717-8781,
Fax: 81-22-747-8780)

FEATURES
source Location/Qualifiers
1. .42942
/organism="Staphylococcus aureus temperate phage phiSL1"
/viralon
/db_xref="taxon:130478"
/note="prophage"
1. .29
/note="core sequence of attL site"
complement(113..1318)
/note="orf401"
phi L54a integrase homologue"
/codon_start=1
/transl_table=11
/product="integrase"
/protein_id="BAB21695.1"
/db_xref="GI:12697823"
/translation="MWIEKFKNNKNETRYRYEYKYPYTDKWKRVSVVLNKNTKQSQ
KAMFRLEEKIEKLNKSSSEKLTTFHALDEWLEYHKTSGSKLTLLNNIKIR
NIKRYSSENLLKLDTRYMQIFINKLSDIYSONVTRQLGDMGAIRYAVKFNYPN
EYLLNVKIPKRRKTIEDIEKDESKWYNLEMNQVLRDHLNKLHKNRILIAS
ILLEVALTRMIGLQALQEKDIDLNLKNTINTGTIRIKYEEGFGYKDTTKTISKR
SISINRVEIEPKLILENKKMLKRNSSYVDGFIPTTKGNPLCNQIAGVLKKTIK
ALNMNKKVTHTFTRHTITLLVMNVSLKAMKRVGHVDEKTIIRIYTHVTEKMREL
TOKLENIPS"
1444..2058
/note="orf204"
/codon_start=1
/transl_table=11
/product="phi PVL ORF 30 homologue"
/protein_id="BAB21696.1"
/db_xref="GI:12697824"
/translation="MTQFLGALLGVGLGYIPYKYLTMIGLVSEKKNKVINTPVLIIIS
IFTCILWFYFTIENNVDLNLNLIQLTGLKANILFLFVLTVFVENPLIVKFIW
LNIITRKMKLDCISLDKDKLNNKGPFIYIKDENRIIEGELKTYNSGDSF
DLLEVERQFKVSDLPNSDNELIKHTLVDLKKQIKLDLYLMNEY"
complement(2291..2686)
/note="orf131a"
/codon_start=1
/transl_table=11
/protein_id="BAB21697.1"
/db_xref="GI:12697825"
/translation="WETNKTIDLMNVFPPKRYTEAKGLVAQPIITTTINSARRVENED
MTVCYLQDDDDVMDIFEDRDIITVYCPCENSGTATDEYFCEIIFNSDDTFTLKRLSNY
TIKDRSPMSKINDVNIITGVVRLFRDFK"
complement(2715..3149)
/note="orf144"
/codon_start=1
/transl_table=11
/protein_id="BAB21698.1"
/db_xref="GI:12697826"
/translation="MKKVIIGLLVSTLALTACGEKEKPKKEKNKKSQTKHKHDSKPKT
QGEKMKVEDKNPPNNSIQNNSNONQNNLNNSDPSNNTPANINENDSQNTNLN
DEYVSPGTRDEQAKAFEEYKKGKEBARAGASAVPGANIN"
complement(3167..3628)
/note="orf153"
/codon_start=1
/transl_table=11
/protein_id="BAB21699.1"
/db_xref="GI:12697827"
/translation="MGLXEETLIQHDYIEIRADVLPNLDGVWGLDILIKRGLSDR
EKAGILFEALNKLTYGDIADYSKFNRRKFENYARRHGFTSAVPLRIVEAYNYGVR
NIYELSEYLIQSEEYILRAIEQYKKIYGTGTHGEYSITFEPLRVFKYKEI"
complement(3641..3955)
/note="orf104a"
putative"
/codon_start=1
/transl_table=11

CDS
/product="repressor"
/protein_id="BAB21700.1"
/db_xref="GI:12697828"
/translation="MCFSKRMKQSKREKOGMTLAEILGRKIGTKTEATVORYESGNIKNLK
NDTIESIATALNVNPAYLMGWVEENDEVOHRAAHLEGEELTDDEWQRLVDYADYIRSK
RK"
4357..4905
/note="orf182"
putative"
/codon_start=1
/transl_table=11
/product="antirepressor"
/protein_id="BAB21701.1"
/db_xref="GI:12697829"
/translation="MQALQIVBNETHYVDSREVAEMIGKRDHNLVRDLKGYIKVLED
SSKLSHNFESTYVNSQNKVQPCYLLTKKGDIVANKMTGSKGTLFTATYVDAPHK
MDEYTRQAQLNVPQTPKCALEMMFAKQDQFNKMQOEITGIRHIVGIETKNRN
DTNKMLSAIAQYFRWRNAPES"
4889..5134
/note="orf81a"
/codon_start=1
/transl_table=11
/protein_id="BAB21702.1"
/db_xref="GI:12697830"
/translation="MHQVKYSEAYKALEEKRCNLKIRMQNRKGMKMLANGATKTQINK
LSKLDVITDEPRLVEIYISVIKSMAIKYGVDISQFEI"
complement(5296..5505)
/note="orf69"
/codon_start=1
/transl_table=11
/protein_id="BAB21703.1"
/db_xref="GI:12697831"
/translation="MKACLYLSNDFVEIDNLEKVIKSGHRTVEISKEIKSSLFTN
GSYTFVGDKIVAIASAKIEFIEFID"
complement(5776..6141)
/note="orf121"
/codon_start=1
/transl_table=11
/protein_id="BAB21704.1"
/db_xref="GI:12697832"
/translation="MPLISDEFDTLTKDOOYILSVLYKDYLECVCVLGSKVLCNPFGS
AKDIIHKYFKLHFEDVKYDLNKLKNSGFLNGVYASNTIYHYTISDKTVVYFENEFKN
NLKSIIDSISKTASIIPLG"
6196..6411
/note="orf71"
/codon_start=1
/transl_table=11
/protein_id="BAB21705.1"
/db_xref="GI:12697833"
/translation="MNTREGLRIGVPOVQVSSKADASSSYLTEKERNLGAELILELIKKS
DYSYLEINKVFYALDRELQYRANNKL"
6436..6699
/note="orf87"
/codon_start=1
/transl_table=11
/product="phi PVL ORF 37 homologue"
/protein_id="BAB21706.1"
/db_xref="GI:12697834"
/translation="MPKIIIPPTPENTYRGEKFKVKKLYATPTQIHLPGVCRCSTGYN
WVKYREDNLGVENLYIDYSATGTLINISKEEYLIRKHKWY"
6712..6873
/note="orf53"
/codon_start=1
/transl_table=11
/product="phi PVL ORF 38 homologue"
/protein_id="BAB21707.1"
/db_xref="GI:12697835"
/translation="MSDITYKSYLLAVLCFTALAIVLMPFLYTTAWSIAGFASIIATFI
FYKEYFYEE"
6966..7268
/note="orf100a"
/codon_start=1
/transl_table=11

CDS
/product="phi PVL ORF 37 homologue"
/protein_id="BAB21706.1"
/db_xref="GI:12697834"
/translation="MPKIIIPPTPENTYRGEKFKVKKLYATPTQIHLPGVCRCSTGYN
WVKYREDNLGVENLYIDYSATGTLINISKEEYLIRKHKWY"
6712..6873
/note="orf53"
/codon_start=1
/transl_table=11
/product="phi PVL ORF 38 homologue"
/protein_id="BAB21707.1"
/db_xref="GI:12697835"
/translation="MSDITYKSYLLAVLCFTALAIVLMPFLYTTAWSIAGFASIIATFI
FYKEYFYEE"
6966..7268
/note="orf100a"
/codon_start=1
/transl_table=11

CDS
/product="phi PVL ORF 37 homologue"
/protein_id="BAB21706.1"
/db_xref="GI:12697834"
/translation="MPKIIIPPTPENTYRGEKFKVKKLYATPTQIHLPGVCRCSTGYN
WVKYREDNLGVENLYIDYSATGTLINISKEEYLIRKHKWY"
6712..6873
/note="orf53"
/codon_start=1
/transl_table=11
/product="phi PVL ORF 38 homologue"
/protein_id="BAB21707.1"
/db_xref="GI:12697835"
/translation="MSDITYKSYLLAVLCFTALAIVLMPFLYTTAWSIAGFASIIATFI
FYKEYFYEE"
6966..7268
/note="orf100a"
/codon_start=1
/transl_table=11

```
/protein_id="BAB21708.1"
/db_xref="GI:12697836"
/translation="MTKNYKDMTQDEIKOLLSEKSGELYELAKEIGESKFOILLFSS
IGVIDGDLAGNSVIGHTFDLASLDSTKSYKDI NVNLMQCKSQNFGLDGDDKED"
7273..7533
/note="orf86"
/codon_start=1
/transl_table=11
/product="phi PVL ORF 39 homologue"
/protein_id="BAB21709.1"
/db_xref="GI:12697837"
/translation="MYEIGDIIIRKNIHVNGDFDKFILKGHMGISIQVDMNVPIK
HAYVDENOLDWASDLFNQDAIDWIEENTDEQRLNLNVKWK"
7546..8082
/note="orf178"
/codon_start=1
/transl_table=11
/protein_id="BAB21710.1"
/db_xref="GI:12697838"
/translation="MNEQLRELETFEQDDRFKVTDLDSANVVKKLDAITTKENIN
ELANKIEIRNEMKDEKVEKLOSGEYLGSLVIEYRIQEQDSKFKLNPYCKVTAR
KGSVKIQVNEQEVILKQLRGFDNVKVKLSQSDIKKDFNVTEGTLDANGVEL
EGASIVKEKPTSVTKVGE"
8083..8733
/note="orf216"
/codon_start=1
/transl_table=11
/protein_id="BAB21711.1"
/db_xref="GI:12697839"
/translation="MAEQNLQKVIADVKANITAGFTKQTKGYNFVSQSIHLHIRE
KMEHNLVLPNTLNWTHFTFKNKQGEVTEFIVEMDLNNTWINDRPEQYEVSY
HAYQNDLSQAHGTALTYAERYFLMKFFNPTDEADAKOKRYSTVSQEFKDL
TREVNDFAIAIKESGFAERYQEQINKLEKNMVEALNKQINVTROQIKKWLGGIEQ"
8730..9173
/note="orf147
putative"

Query Match          94.0%; Score 149.4; DB 7; Length 42942;
Best Local Similarity 96.2%; Pred. No. 5.3e-35;
Matches 153; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 atggtaccagaagtattttaaactaaactagtggttcagatgtagctgcagaa 60
- Db 11898 ATGCTACCAAGAATTTTGAATAATTAACCTTTCAGTGTTCAGATATGACGCTCAGAA 11957
|||||
QY 61 ctcatagatgagcacagcgcatgaaataggtgttacgacctattatccaaaactt 120
|||||
Db 11958 CTATAGACGAGGCACAGGCGGATGAATAAGTTATATGACCTATTATCCAAAAACTT 12017
|||||
QY 121 gcaaacgtcatcacagcccgctatctgcgaattataa 159
|||||
Db 12018 GCAGAGGTCATACACGCCCGCTATCGTCGAATATTA 12056
|||||

RESULT 2
AP003360
LOCUS
DEFINITION
Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete
sequence, section 3/9.
ACCESSION
AP003360 BA000017
VERSION
AP003360.2 GI:14246388
KEYWORDS
SOURCE
Staphylococcus aureus subsp. aureus Mu50 (sub_species:aureus Mu50,
strain:Mu50) DNA.
ORGANISM
Staphylococcus aureus subsp. aureus Mu50
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE
1 (sites)
AUTHORS
Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I.,
Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Lian, J., Ito, T., Kanamori, M.,
Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A.,
Mizutani-U, Y., Kobayashi, N., Tanaka, T., Sawano, T., Inoue, R.,
Kaito, C., Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S.,
```

```
Yabuzaki, J., Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K.,
Yoshino, C., Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H. and
Hiramatsu, K.
Whole genome sequencing of methicillin-resistant Staphylococcus
aureus
Lancet 357 (9264), 1225-1240 (2001)
MEDLINE
21311952
REFERENCE
2 (bases 1 to 348527)
Ohta, T.
Direct Submission
Submitted (28-FEB-2001) Toshiko Ohta, University of Tsukuba College
of Medical Technology and Nursing, Department of Medical
Technology; 1-1-1 Ten-nodai, Tsukuba, Ibaraki 305-8577, Japan
(E-mail: tohta@tsakura.cc.tsukuba.ac.jp, Tel: 81-298-53-3454,
Fax: 81-298-53-3454)
On May 29, 2001 this sequence version replaced gi:13874937.
FEATURES
Location/Qualifiers
1..348527
/organism="Staphylococcus aureus subsp. aureus Mu50"
/strain="Mu50"
/sub_species="aureus Mu50"
/db_xref="taxon:158878"
158..718
/gene="SAV0620"
158..718
/gene="SAV0620"
/notes="similar to F1mE recombinase"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB56782.1"
/db_xref="GI:14246389"
/translation="MNKVEAKFNDIVKMYEALKIKSERDYLFKKLAIHSGLVSEEL
LFTTVSQVRLIEKCTLSEMKAFHSLIKIRLPETLSKELLOYIEDRSLSNEDVLFO
SLRTNQLSRQOAYRIHQASIEAGIDNVGLTLRTFTFAYHAYKQGIPIPIQKYLGH
QSAIETLNFIGLENECEHSIYISLQL"
737..2761
/gene="SAV0621"
737..2761
/gene="SAV0621"
/notes="similar to NADH dehydrogenase"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB56783.1"
/db_xref="GI:14246390"
/translation="MSLVYLLIALLVIMAMILLMSKRRALAKYAGYIALVAPVISSYI
FLIQIPSAKQYLSIPWIKRTIDINLDLGLSLMFLSLIISLIGIAVFFYATQYL
SSRKNDLPFRFYFLTFMFMSMIGIVISDNTILMYIFWELTSVSSFLISYWNQDSQ
FGAIQSFMTYVFGGLALLVGFIMLYIMTGTNNITDILGQADHIKHNGLFIPMIFMFL
GAFTKSAQFPFHWLPRMAAPTVPVSAYLHSAVMKAGIFLLRFTPLGLSNMYIYI
VTFVGLITMLGSIATALKOWDLGILAYSTISOLGIMAMVGGIGGYAOHQODALYAI
YVFLGALFHLNMHAIKCALPFGVGLDHEAGSDIRILSCMRQLFPMKNLVMTIA
ALSMAGVPEFLNGELSKEMFDALTQGLSQFSLISMAIVFVGVIASIFFTVLYM
VKEVFTKYDSKVFTKNIHEPWLFSPLILMVLVPVFFVFNIFKGIIVIPALRGV
SGNHQIDPLAHVQSHQGNFIPLLITIIILIGLSLAIKVMKKVFTGKIQTIVSK
GYEMVYRHFKEKATKFRKRVMDRLNQYIIMTIGIEMIIGYGVIRIGLQPKVHQLHVS
EFGALEIILAIYTVTIGISILIFIRQLRTWILNGVIGFVVTLFFIAKMAP"
2807..3139
/gene="SAV0622"
2807..3139
/gene="SAV0622"
/notes="similar to NADH dehydrogenase"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB56784.1"
/db_xref="GI:14246391"
/translation="MYCQFRLNPNRNSNANKREIKISVSLMALIVVSLIFITQO
TDGLSISISDFYKADKLTKGKNIVNALIGDFRALDTLFEGLVLIITGLGIYTLINQD
RKQGDRE"
3126..3551
/gene="SAV0623"
```


REFERENCE 2 (bases 1 to 43081)
AUTHORS Sugai, M., Yamaguchi, T., Hayashi, T., Nakasone, K. and Takami, H.
TITLE Direct Submission
JOURNAL Microbiology: Kasumi 1-2-3, Hiroshima University
Faculty of Dentistry, Microbiology: Kasumi 1-2-3, Hiroshima,
Hiroshima 734-8553, Japan (E-mail: sugai@hiroshima-u.ac.jp,
Tel: 81-82-257-5637, Fax: 81-82-257-5639)

FEATURES
source
1..43081
/organism="bacteriophage phi E7A"
/specific_host="Staphylococcus aureus E-1"
/db_xref="taxon:106284"
complement(53..1102)
/note="ORF1
similar to phage phi 11 integrase"
/codon_start=1
/transl_table=11
/protein_id="BAA97587.1"
/db_xref="GI:8918416"
/translation="MPVYKNDWTGWYFSIRKYDVYGNKKRMQGFSTTKREAKRAEA
IFLNDVNEGYSDSKTFDYFHHYLENSDLRPKTKRKQNEYHKHFKAFGHKMKKIT
QNOCDFKYLTIENVASTNSARTIWSGFKVINYAKYFGLRDTPTISIKPIPRVKPK
PKFMREFEKRIKDIEQDYRELFTLMFYTLGLRIGEAALVWTDYNYKKEISINKT
KDISNRTIYPRKTSSEDIYPLPKFINTMLTERHQREKELNKYFDERSYFIFGGMAP
KHYSVQKKQKAFPHYNIHALRHSIYSTLANNGVDIYVQLSLMRHAQITETMGTYSH
LYTKKKDAIAIFDK"
complement(1160..1531)
/note="ORF2
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAA97588.1"
/db_xref="GI:8918417"
/translation="MENKHYHTDAARPNOVVDYKYESLLNIRONKKMEKGNYS
GGGKNGKVIYKEVEQEKRIDNKLDSLPRMDKMDAKISGLEARQTKWFGVIAIS
TAIGLSLIVAGLVNLFY"
complement(1521..1946)
/note="ORF3
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAA97589.1"
/db_xref="GI:8918418"
/translation="MTNIDWITSNRVFNPNNSNLLIDEPITTSISLFWNPQYTTTIS
FGISNIDNDVQITNLRIGNDLDTKEIVNAPLYIDAKENEVKNKRANGLVDVSTVT
LNNFLFEETGNHYITPKLGESTQTCYFNVLNQLEDGK"
complement(2002..2187)
/note="ORF4
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAA97590.1"
/db_xref="GI:8918419"
/translation="MSTYKEIEHLHINTGGKELTQEQIEEAKSFIDSOGFKMDIREAK
ESRQRVMESKITDRTKM"
complement(2184..2330)
/note="ORF5
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAA97591.1"
/db_xref="GI:8918420"
/translation="MNLKEVDINIEEMVEIPFYTEBELTYRLNGLPLITKSELEEQ
ESKK"
complement(2342..3058)
/note="ORF6
similar to phage phi PVL repressor"
/codon_start=1
/transl_table=11
/protein_id="BAA97592.1"
/db_xref="GI:8918421"
/translation="MNSFKDLKQIMSERKTSQSLSRRTGIGRNSISDYLNKGYEAK
QDKVFELAKVLNVNEAWLWGFDISKNRKIENNDITSYNNKLTPTPPQNVLNAYNSQLD

CDS
BONSKGDNIVDINSYKQDKTPVNVNGCVSAGVGERLHDETLETEMVKAPVPPHDLALK
VNGDSMEPMFKDGEILIFVEKTHNIRNGQIGIFIEEEATVVKVFEVDDRLTLVLSLNNK
3222..3464
/note="ORF7
similar to phage O1205 ORF5"
/codon_start=1
/transl_table=11
/protein_id="BAA97593.1"
/db_xref="GI:8918422"
/translation="MIYNFDYSLLYERMAERYSSOSSLANAIPISRTSINHKLQGNL
FTQWEIKRICELLEIPPTKVGRYFFEQNVQKTVQTS"
3480..3659
/note="ORF8
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAA97594.1"
/db_xref="GI:8918423"
/translation="MEETIKHLEFRKQFTPAQWHEINRIIDGQFSKKAELRLDDQD
VEVIKNIITQQKIMK"
complement(3760..4242)
/note="ORF9
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAA97595.1"
/db_xref="GI:8918424"
/translation="MEFIQSTLFSNVVAFALGLSAYSIFYTRSONKFSFVLSDLNFI
YENNEVELNFVANDSSRTHLEELIFLDKKNVLTPTINVLESSEYSLGIYNPSYL
HAPIDKQDKPEVMIANSSSEFLYKFELEPAFIVKISNORINKLKKYKILISTDSVEHN
"
4687..4872
/note="ORF10
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAA97596.1"
/db_xref="GI:8918425"
/translation="MLNLKELREKGIYTRQLAKLTELQNSTIRSIETEVKNPGLTV
KKICDALQVDIANVKEK"
4874..5626
/note="ORF11
similar to phage phi PVL antirepressor"
/codon_start=1
/transl_table=11
/protein_id="BAA97597.1"
/db_xref="GI:8918426"
/translation="MQALQTKSSIGEMFNIEQKENGETAISGRELHQALEVKTAYKDW
FPRMLKYGFENTDYTAIAOKRATAGQNMTHYIDHALTDTAKEIAMTQISPEPKRAR
QYFIOVERAWNSPEMIMORALAIANTINOLETKIERDKPKIVFADAVATTTSILVG
ELAKTIKONGINIGORRLFEWLRONGFLIKRKGVDYNMPTQYSMERELFEIKETSITH
SDGHTSISKTPKVTGKGQOYFVNKFLGETQT"
complement(5731..6396)
/note="ORF12
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAA97598.1"
/db_xref="GI:8918427"
/translation="MNNINFTQRLDLTKKNOAILSKLPVEAYAKAANTMNNYSYVYNA
LEIQTSVNNMNSIRINQSKLSDMASYHQVTKNHPMFKSNLFSKILDEFISSNSFP
DDEQKTSVTHLRKSFVDVDPVLGKTVNSAHPIDDDVNTKESDKIFYKSIINOYFLAPS
SSFVHDVSLTVAKSVAVNMVVRTANDDYVNYFFSTAVIAVCYVASCANAFDLKNNRK
DFK"
6467..6688
/note="ORF13
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAA97599.1"
/db_xref="GI:8918428"
/translation="MIKNSLOAKELAVILSVSKSAGQIIRLUNKELEDEGYIAIRGR

```

CDS
  6681..6842
    /note="ORF14
    similar to phage phi PVL ORF38"
    /codon_start=1
    /transl_table=11
    /protein_id="BAA97600.1"
    /db_xref="GI:8918429"
    /translation="MSKTKSYLLAVLCFTVLAIVLMPFLYETTAWSTAGFASITATFI
    FYKEYFVEE"
    6936..7196
    /note="ORF15
    similar to phage phi PVL ORF39"
    /codon_start=1
    /transl_table=11
    /protein_id="BAA97601.1"
    /db_xref="GI:8918430"
    /translation="MYIEIGETIIRKNIHVNGFDKIFILKGHMGISIQVKDMNVPK
    HAYVDENOLDMASDLFNOAIDEWTEENTDEQRLINLVNR"
    7206..7427
    /note="ORF16
    unknown"
    /codon_start=1
    /transl_table=11
    /protein_id="BAA97602.1"
    /db_xref="GI:8918431"
    /translation="MNQTVYIIRHRDMPYITNKPTDNNSDYSYSTNRNRAREFNGM
    EPASINMDYHKAIRKATVTETIEYEVEHD"
    7420..8199
    /note="ORF17
    unknown"
    /codon_start=1
    /transl_table=11
    /protein_id="BAA97603.1"
    /db_xref="GI:8918432"
    /translation="MTQQTNDQVDILTGLGVKDISKONANKYKFAIYKGFQGTGTF
    LTKNNALVIDINDEGTVTVEDGAVQLKHYKFSAVIKMLPKIIEQLRENGKQIDVL
    VIETIQLRDITDDINDGSKRPTFNDGECATRVISYIRYSIKLQHYOFHLAIS

Query Match      33.18; Score 52.6; DB 7; Length 43081;
Best Local Similarity 58.7%; Pred. No. 1.9e-05;
Matches 91; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 5 taacaaagaatttttaaaactaaacttgagttgcagatgtagctgcgaactca 64
Db 11922 TCACAAACAAATATAAAAGAAATATTACATTGTAGACATGATATGCTCAAAAGATGA 11981
QY 65 tagatgagggcacagggcgatgaaaataggtgtacgacctatttatccaaaactgcag 124
Db 11982 TTGATTTTGCNAACGGAGACCAAGAACTTAAAAAACTTATTGATGATAGTTGAAG 12041
QY 125 aacgtcatcacgcccgcgtatcgtcgaaatataa 159
Db 12042 AAAAAGAAGAAGACCGCAATCGTCGATATATA 12076

RESULT 4
AB044554
LOCUS
DEFINITION
  Staphylococcus aureus prophage phipV83 proviral DNA, complete
  sequence.
ACCESSION
  AB044554
VERSION
  AB044554.1
KEYWORDS
  GI:8918747
SOURCE
  Staphylococcus aureus prophage phipV83 (strain:P83,
  specific_host:Staphylococcus aureus) proviral DNA.
  Staphylococcus aureus prophage phipV83
  Viruses.
  1 (sites)
REFERENCE
  Kaneko,J., Kimura,T., Kawakami,Y., Tomita,T. and Kamio,Y.
  Pantom-valentine leukocidin genes in a phage-like particle isolated
  from mitomycin C-treated Staphylococcus aureus V8 (ATCC 49775)
  Biosci. Biotechnol. Biochem. 61 (11), 1960-1962 (1997)
MEDLINE
  98067870
  2 (sites)
  Zou,D., Kaneko,J., Narita,S. and Kamio,Y.
  Complete nucleotide sequence and molecular characterization of
  prophage PV83pro carrying lukM-lukF-pv(p83) gene cluster in
  Staphylococcus aureus strain P83
  Unpublished (2000)
JOURNAL
  3 (bases 1 to 45636)
  Kaneko,J., Zou,D. and Kamio,Y.
  Direct Submission
  Submitted (09-JUN-2000) to the DDBJ/EMBL/GenBank databases. Jun
  1-1 Tsutsumi-dori Amamiyamachi, Aoba-ku, Sendai, Miyagi 981-8555,
  Japan (E-mail: jkaneko@biochem.tohoku.ac.jp, Tel:81-22-717-8781,
  Fax:81-22-717-8780)
  Location/Qualifiers
  1..45636
    /organism="Staphylococcus aureus prophage phipV83"
    /proviral
    /specific_host="Staphylococcus aureus"
    /strain="P83"
    /db_xref="taxon:129009"
    1..156
    /note="attP region of phi 11"
    complement(423..1139)
    /note="C-terminal portion truncated
    orf 1"
    /codon_start=1
    /transl_table=11
    /product="integrase"
    /protein_id="BAA97808.1"
    /db_xref="GI:8918748"
    /translation="MEVYKDDNTGKWFYSIRYKDYVGNKKMKRGFKKDAKLAES
    EFTONVYKGSNDQPEYIFFDLKNNLSARSIEKRTTEYNTHIKERFGNIPICKIT
    TQCTATFREEDFKEVEQITNDTSRQLTKLLFYSLRIGELALOWKDYNKIKGEIDVNF
    KINLSNRKIEYNKKKALKG"
    1252..1866
    /note="orf 2"
    /codon_start=1
    /transl_table=11
    /product="phi PVL ORF 30 homologue"
    /protein_id="BAA97809.1"
    /db_xref="GI:8918749"
    /translation="MTQFLGALLTGVLGYIPYKYLTMIGLVSEKNIINTPVLIFS
    IETCLIFWFTFIENNVDLKNLSLLOLLTGLKANIMELIIFLVTLVFNPLVFKFTIW
    LINKTRKFNWLDICISLDKRLKLNNGKPVFIKDFENRIIEGELKTYNSAGSDF
    DLLEVERODFKVSDLPSENDELYIKHTLVLDLKQQIKLDLYLMNEY"
    complement(2098..2547)
    /note="orf 3"
    /codon_start=1
    /transl_table=11
    /protein_id="BAA97810.1"
    /db_xref="GI:8918750"
    /translation="MEKQYLWRYNDIEKRTNMTNKTIDLMNMYVEFPKRYTEAKGL
    VAQPTTINSTRRVENGMTVCYILDQDDVIDIFIDRDIITVFFPENGTETDEYFCE
    ILFNSDDTFLKRLSNYTKDORSPMSKINDVNTGKVVLFRDFK"
    complement(2583..2768)
    /note="orf 4"
    /codon_start=1
    /transl_table=11
    /protein_id="BAA97811.1"
    /db_xref="GI:8918751"
    /translation="MSSYKEIEHLHINTGGKELTQDIEAKAFIDSQEFKDMIREAK
    ESKQVWESKITDRTKM"
    complement(2923..3633)
    /note="orf 5
    putative"
    /codon_start=1
    /transl_table=11
    /product="repressor"
    /protein_id="BAA97812.1"
    /db_xref="GI:8918752"
  
```


code for large proteins in vitro
DNA Res. 7 (1), 65-73 (2000)
20181126
2 (bases 1 to 4620)
Ohara, O., Nagase, T. and Kikuno, R.
Direct Submission
Submitted (31-JAN-2000) to the DDBJ/EMBL/GenBank databases. Osamu
Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology:
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:cdnaif@kazusa.or.jp, URL: http://www.kazusa.or.jp/huge/,
Tel:+81-438-52-3913, Fax:+81-438-52-3914)

FEATURES

source

1. .4620
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="pBluescriptII SK plus"
/tissue_type="brain"
141. .3158
/gene="KIAA1361"
141. .3158
/gene="KIAA1361"
/note="Start codon is not identified."
/codon_start=1
/product="KIAA1361 protein"
/protein_id="BAA92599.1"
/db_xref="GI:7243103"
/translation="LLSRMPSTNAGSLKDPETAEALFFKEDPEKLTDLREIGHGSGF
AYFARDVRTNEVVAIKKMSYSGKSTKQDIIKEVKFLORIKHNSIEYKGCYLRE
HTAWLMEYCLGASDILEVHKPLQREVEAATHGALGLAYLHSTWTHRIDKAGN
ILLTEPGQVCLADPGSASMASPANSFVGTPTWAPDEVILAMDEQYDGVDMVSLGIT
CIELAERKPLFNAMASALYHIAQNSPTLQSNWSDYFRNVDSCLOKIPQDRPTS
BELLHVLIRERETVLDLQRTDAVRELDNQLYRKMKLLFOEAHNGPAVEAQE
EEEDHVGRTGTGVNSGNSQIPSMISASOSSSVNSLDPVSDKSELDMEGDH
TVMSNSVILHKPEENYREGDPTRASDPQSPPOVS8HKSHYRNRHFAITRTASL
VTROMQHEODSELREQMSGYKRMROHQQLMTLENKLAEMDEHRLDKDLETQR
NFEAENEKLIKHOAMEKAEKVMNEKKFOOHLQAOQKELNSFLSQREYKLR
KEOLKELENQSTPKKEOWLSKQENIQHOAEANLLRQRYLELCRRFKR
RMLGRNLGSDLVRELNKRQTKDLEHMLLRQHESMQELFRHLNTIQKMRCELI
RLQHTSLLTNOLEYKREKRLKRVHMEVROQPSLKSKEIQIKKQFOODTCKIOTRO
YKALRNHLETPPKSEHKAVKRLKEQTRKLAILEAOYDHSINEMLSQALRLLEE
EACQVLMQLOLELLNAYQSKIKWQAEAOHDRELLEQVRSILRRALLEQLEEE
MLALQNTERTIRSLLEQRAEIEAFDESMBRGLNSVLSLSEARSHSTPGASG
SHNPTGPGHPGHPGHPGHPGHPGHPGHPGHPGHPGHPGHPGHPGHPGHPGHPG
LRRTASGRTGECMSRSTVTSQISNGSHWSYT"

BASE COUNT 1490 a 942 c 1039 g 1149 t
ORIGIN

Query Match 23.1%; Score 36.8; DB 9; Length 4620;
Best Local Similarity 58.0%; Pred. No. 0.91;
Matches 65; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 15 atttttaaaactaaacttgatttcagatgatgtacgctcagaaactcatagatgagc 74
Dy 3392 ATTAAATAAATGTGTATGTGTATATATATATATATATATATATATATATATGC 3451
Qy 75 acaggcgatgaaataggcttgacacacctatttccaaaacttcagaa 126
Dy 3452 ATGTGGTGAAGAAATGGCTAGTAGGAGGATTTTCTGAACACTGCAAAA 3503

RESULT 6

CEC38C6

LOCUS

DEFINITION

Caenorhabditis elegans cosmid C38C6, complete sequence.

ACCESSION

293375.1

VERSION

GI:1914028

KEYWORDS

HTG.

SOURCE

Caenorhabditis elegans.

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;

Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REMARK

REFERENCE

AUTHORS

TITLE

JOURNAL

1 (bases 1 to 25943)
none

Genome sequence of the nematode *C. elegans*: a platform for
investigating biology. The *C. elegans* Sequencing Consortium
Science 282 (3596), 2012-2018 (1998)
99069613

The *C. elegans* Sequencing Consortium.

2 (bases 1 to 25943)

Baynes, C.

Direct Submission

Submitted (27-MAR-1997) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or tw@nematode.wustl.edu

Coding sequences below are predicted from computer analysis, using
predictions from Genefinder (P. Green, U. Washington), and other
available information.

Current sequence finishing criteria for the *C. elegans* genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is not the entire insert of clone C38C6.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.

The true left end of clone C38C6 is at 31320 in

sequence 293378.

The true left end of clone ZC101 is at 25843 in this sequence. The
true right end of clone F9H8 is at 6753 in this sequence. The
start of this sequence (1..104) overlaps with the end of sequence
293378.

The end of this sequence (25843..25943) overlaps with the start of
sequence 293395.

For a graphical representation of this sequence and its analysis
see: [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=C38C6)
name=C38C6.

FEATURES

source

1. .25943
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="II"
/clone="C38C6"
complement(join(2285..2341,2388..2546,2592..2646,
2691..2947,3118..3191,3238..3532,3750..3856,3973..4203,
4253..4388))
/gene="C38C6.5"
complement(join(2285..2341,2388..2546,2592..2646,
2691..2947,3118..3191,3238..3532,3750..3856,3973..4203,
4253..4388))
/gene="C38C6.5"
/codon_start=1
/protein_id="CAB07565.1"
/db_xref="GI:3874865"
/db_xref="SPTREMBL:O45301"
/translation="MLFRTSLVLSALLIIVDAKSGEHLKKVPRIMKEFSTRIF
YEGWLVIQPEKRDWNLHDSIINTKHEGQSLANKMAEASSENFAFFADSKWIE
ALFDDDMVEYCGSEKMKNTQITMLPARESVKVKYKKNMPYAKIMPTKDVLOEH
IKHSAYATHIAQSGSCRDGLGVHVSIESEALKNLDIKYSQOIKMKTFRFHP
ILANSYGPYKMYSPFDPGASIQVCRARETQADYTAEQFISFYRKFLMMWRPKDGD
TDFLDVENEKLFITNLEWFCPTVDLMDLSLIAREVVAEKLDEVAADYPTTFEFS
SVDFLNRTKLNLLKTESCLGSMVSDHFRSRSRSGSGSTNSYLSVSYRVOYE
DIVLPAEAKTHFMQTHAGDFMDPVFHTVWGKRDYEHKWFENLWDFDQFYVQSL
KVEAHR"

gene

CDS

gene

CDS


```

TITLE
JOURNAL
COMMENT
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (04-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 30, 2001 this sequence version replaced gi:14010896.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Center project name: L8841
Center clone name: 659_G_19
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 8823: contig of 8823 bp in length
* 8824 8923: gap of 100 bp
* 8924 25814: contig of 16891 bp in length
* 25815 25914: gap of 100 bp
* 25915 63597: contig of 37683 bp in length
* 63598 63697: gap of 100 bp
* 63698 145632: contig of 81935 bp in length
* 145633 145732: gap of 100 bp
* 145733 153460: contig of 7728 bp in length
* 153461 153560: gap of 100 bp
* 153561 175931: contig of 22371 bp in length
* 175932 176031: gap of 100 bp
* 176032 187114: contig of 11083 bp in length
* 187115 187214: gap of 100 bp
* 187215 189009: contig of 1795 bp in length.
*
* Location/Qualifiers
* 1..189009
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="17"
* /map="17"
* /clone="RP11-659G19"
* /clone_lib="RPCI-11 Human Male BAC"
*
BASE COUNT 51947 a 39987 c 40656 g 55623 t 796 others
ORIGIN

Query Match      23.1%; Score 36.8; DB 2; Length 189009;
Best Local Similarity 58.0%; Pred. No. 1.6;
Matches 65; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 15 atttttaaaactaaactaggttcagatgatgcgtcagaactcatagatgagc 74
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36136 ATTTAAATAATATGCTATGTGTGTCATATATATACACACACATATATTATGC 36077
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 75 acagggcgatgaataatagggtgtacgacatttatccaaaactgcagaa 126
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36076 ATGTGGTGAAGAATTTGGCTAGATAGGGGATTTTCTGACACACTGCAAAA 36025
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
AC023181
LOCUS
DEFINITION
AC023181 149109 bp DNA HTG 04-JUL-2001
Homo sapiens clone RP11-20D20, WORKING DRAFT SEQUENCE, 5 unordered
pieces.
ACCESSION
AC023181
VERSION
AC023181.4 GI:14595905
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

```

```

SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 149109)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-20D20
unpublished
2 (bases 1 to 149109)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bida,F.,
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 4, 2001 this sequence version replaced gi:12039539.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Center project name: L3882
Center clone name: 20_D_20
-----
Summary Statistics
Sequencing vector: M13; M77815; 3% of reads
Sequencing vector: Plasmid; n/a; 97% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 147126 bases at least Q40
Consensus quality: 148104 bases at least Q30
Consensus quality: 148478 bases at least Q20
Insert size: 150000; agarose-fp
Insert size: 148709; sum-of-contigs
Quality coverage: 8.9 in Q20 bases; agarose-fp
Quality coverage: 9.0 in Q20 bas.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 5226: contig of 5226 bp in length
* 5227 5326: gap of 100 bp
* 5327 17559: contig of 12233 bp in length
* 17560 17659: gap of 100 bp
* 17660 74088: contig of 56429 bp in length
* 74089 74188: gap of 100 bp
* 74189 115843: contig of 41655 bp in length
* 115844 115943: gap of 100 bp
* 115944 149109: contig of 33166 bp in length.
*
* Location/Qualifiers
* 1..149109
* /organism="Homo sapiens"

```

FEATURES
source

[illegible][illegible]

Consensus quality: 226966 bases at least Q20
 Insert size: 222542; sum-of-contigs
 Quality coverage: 10.14x in Q20 bases; sum-of-contigs

```

FEATURES
  Source
    Location/Qualifiers
      1..222542
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="3"
        /map="3p"
        /clone="996C6"
BASE COUNT      57203 a 54087 c 54080 g 57172 t
ORIGIN
  1..222542
    22..5%   Score 35.8; DB 9; Length 222542;
    Best Local Similarity 63.2%; Pred. No. 3.3;
    Matches 55; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
  QY 72 ggcacggcgatgaataagttgtacgacctattatccaaaaacttcgagaacgtca 131
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  Db 5160 GACGAGGAGCAATAAAAAATCATTTGTTGTGCAAAAGTGTCCAAAAGTTCACAGAGCTA 5101
  QY 132 tacacgcccgcctatcgtaattatta 158
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  Db 5100 TACATGCCCCAGGCTGTAGGCAACAATTA 5074

```

```

RESULT 15
AC006220 LOCUS AC006220 105222 bp DNA PLN 05-APR-2000
DEFINITION Arabidopsis thaliana chromosome II section 28 of 255 of the
complete sequence. Sequence from clones F16J10, T20G20, T3P4.
ACCESSION AC006220 AB02093
VERSION AC006220.3 GI:6598532
KEYWORDS HTG;
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 105222)
Lin.X., Kaul, S., Rounsley, S.D., Shear, T.P., Benito, M.-I., Town, C.D.,
Fuji, C.Y., Mason, T.M., Bowman, C.L., Barnstead, M.E.,
Feldblyum, T.V., Buell, C.R., Ketchum, K.A., Lee, J.J., Ronning, C.M.,
Koo, H., Moffat, K.S., Cronin, L.A., Shen, M., VanAken, S.E., Umayam, L.,
Tallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H.,
Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D.,
Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and
Venter, J.C.
Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana
Nature 402 (6763), 761-768 (1999)
20083487
10617197
2 (bases 1 to 105222)
Lin.X.
Direct Submission
Submitted (09-MAR-2000) The Institute for Genomic Research; 9712
Medical Center Dr., Rockville, MD 20850, USA
On Dec 17, 1999 this sequence version replaced gi:4581161.
The sequence and annotation of chromosome 2 were merged from those
of the individual clones on this chromosome after removing
overlaps. For detailed information, please see the TIGR web site
(http://www.tigr.org/tdb/at.html).
Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL
(ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green,
University of Washington), Genscan (Chris Burge,
http://genomic.stanford.edu/GENSCANW.html), and NetPlantGene
(http://www.cbs.dtu.dk/services/NetGene2/), searches of the
complete sequence against a peptide database and plant EST

```

databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone F13P4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

```

FEATURES
  Source
    Location/Qualifiers
      1..105222
        /organism="Arabidopsis thaliana"
        /cultivar="Columbia"
        /db_xref="taxon:3702"
        /chromosome="II"
        complement(<1..1080)
        /note="Sequence from clone F16J10"
        join(<658..774,856..>.1305)
        /gene="At2g05430"
        <658..>.1305
        /gene="At2g05430"
        /note="At2g05430"
        /note="F16J10.2; predicted by genefinder"
        join(658..774,856..1305)
        /gene="At2g05430"
        /note="hypothetical protein"
        /codon_start=1
        /protein_id="AAD24664.2"
        /db_xref="GI:6598533"
        /translation="MSKISEKSTLSHGCTAYDEDLDNDEIVNIKGFWRKSLGLAK
RVFEKHPETTTKFKLSERAKETILTALDLIDIMNLPOOSLFEALKEAENTLIDL
EAAGFKLDWKKRRLKEIRVTRKAKNRRTMRRLDRKIOHLEELSVLQEMKKRQFE
AMCDKPEYLDLSLCLNEVTCDSSEPTK"
        complement(1081..105021)
        /note="Sequence from clone T20G20"
        complement(2070..2185)
        /rpt_family="(TAAA)n"
        2709..2826
        /rpt_family="(TAAA)n"
        2750..2864
        /rpt_family="(TAA)n"
        3314..3367
        /rpt_family="(TA)n"
        complement(4610..4638)
        /rpt_family="POLY-A"
        join(<7338..7419,7778..8372,8698..8803,9111..9231,
9619..9931)
        /gene="At2g05440"
        <7338..9931
        /gene="At2g05440"
        /note="T20G20.21"
        join(7338..7419,7778..8160)
        /gene="At2g05440"
        /codon_start=1
        /product="putative glycine-rich protein"
        /protein_id="AAD24663.1"

```

```
/db_xref="GI:4581180"
/translation="MASKALILGLFSLVLLVSEVSAAROSGMVKPSEETVQPEGYG
GGHGGHGGGGGGHGGHGGHGLDGYGGGGHYGGGGHYGGGGHYGGGGHYGGGGH
YGGGGHYGGGGGGHGGGGHYGGGGHYGGGGHYGGGGHGLNEPVTQKPGV"
7837..7956
/rpt_family="(GGA)n"
7963..8117
/rpt_family="(GGA)n"
/rpt_family="(TA)n"
/rpt_family="(8196..8239)
complement(8196..8239)
join(<10406..10484,10722..11629,11714..11912,11998..12551,
12661..12765,12862..12972,13053..13166,13241..13333,
13479..13556,13638..13739,13811..13888,13964..14032,
14146..14426,14514..14598,14701..14762,14805..15348)
/genes="At2g05450"
<10406..>15348
/genes="At2g05450"
/note="T20G20.20; predicted by genscan"
join(10406..10484,10722..11629,11714..11912,11998..12551,
12661..12765,12862..12972,13053..13166,13241..13333,
13479..13556,13638..13739,13811..13888,13964..14032,
14146..14426,14514..14598,14701..14762,14805..15348)
/genes="At2g05450"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAD24662.1"
/db_xref="GI:4581179"
/translation="MAATKRKRRRGVSDIDEASAAATNOGSTNDPGLPQLFALD
HYPKAINAYSPEYQIAVALKDTPEMTFLDPSFGDLFRPTNKAESGKVLVG
LICHQLVTKRYEMMVFAFHPRLGLREWALITGLEGTYPKNKDVENVMQROEGN
TIMATFGDDKADPTVEELDRILSETDMPAWKKALALIIIVDGLVICDPSPLRPN
EMVELTKNLDFFCKYPLARTSFLTLERIASFGETDVOKLRSGCKQHSYALHGPPL
GLQLAFKLTPTSASLLPSEGDTAIFERTIKNSKLRPIKTSTILACESAEKVDIPY
IMPADSNCPKSEVMPKEDEPRVDNIRMLKSGQEWDEKMWGGDDAPKKOVRPSAT
GERVILRRRVADPVENKKAISHDNGEGSPSGDGEQSGGGGEGGSGDGEV
GDEAFKAYELRRFKQTOLLNKFEDLVFVREIRAAALOSQDNNEEGPRASHKS
DPSFKVEVTKTVKKNKRVKVEKAAEOVQIKSVKSTKQKHVPKRSRLN
NPKKAATGSLPVEEPVNGDGEVSARKHTVSDYDMDNIVILGASSTEGEGRNSIE
EESKLHQNVSQVMDNVDHVPSPHQPOG IPDGNFEEPLPDPILSDPAIEKMSD
DVSSSHQDVAKGDNREEDVAPESQLEEQSPVEVKAETICGDGEALKEAKSPS
VDEALEDTALPGVPPRTGDFDYEARLSDSPPTVVSVKVLTKLDDILAENYSKI
PEKVVVPEVLTQLKDDILEKVESEKVAPEETEGKDDVVEAGVSOEAIIVAPVNEKE
KKSRRRVTFSDDTKNDDKGPDIIVFEEREAGFEVFKAGDSLYNPLEKVEAKFOK
LCNIMOGKRCRCHTLEGRKVGNEFFITLAEANDWGVSYHMMALLMWHKNGEHMKN
RCATAFKSCRNPETFKWDKHLTRPKGECNCEPQLHWIRVDVDFVPMNWKDEKHW
GLAICLERATEIMDPLRSUTRESVVRSLIPIMEMLLYVRATCKOILDAFYPVTPF
TVIRNQLAQNPITTGDCGPYAMEFIELMLNTPVEDMFSIEDDMINFRKGYTDLXE
HGIGSLT"
12065..12163
/rpt_family="(TGG)n"
<15710..>16075
/genes="At2g05460"
<15710..>16075
/genes="At2g05460"
/note="T20G20.19; predicted by genscan"
15710..16075
/genes="At2g05460"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAD24661.1"
/db_xref="GI:4581178"
/translation="MVNPTISDDLLAMAKKVAADKSWLDLKLPIRSGTCGRDVIYRPD
VLKDANIYSLGRVPDDFQVGGGHNPFGQEGRNRPFFKRCLLANNSTVYNEAIRVL
IHETDINGALKLIQRHAPV"
join(<17074..17323,17365..17768)
/genes="At2g05470"
<17074..>17768
/genes="At2g05470"
/note="T20G20.18; predicted by genscan"
join(17074..17323,17365..17768)
/genes="At2g05470"
/note="hypothetical protein"
/codon_start=1
```

```
Query Match      22 48;   Score 35.6; DB 8;   Length 105222;
Best Local Similarity 62.28; Pred. No. 3.4;
Matches 56; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
QY 4 gtaaccaaagaattttcaaaaaactaaacttgagtggtcagatatgtacgctcagaaactc 63
Db 7626 GTTTCGAACTGATATTTAAAAACTATAGTTTCGAGTTTATATATGTTACTTTAAAAAAA 7685
QY 64 atagatgaggcacagggcgatgaaaaataag 93
Db 7686 ATTGAGGTGCAAACTAAATTGAAGTCAGG 7715
```

Search completed: March 1, 2002, 14:13:35
Job time: 1499 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2002, 13:48:36 ; Search time 1172.17 Seconds
(without alignments)
1457.620 Million cell updates/sec

Title: US-09-407-804A-8

Perfect score: 159

Sequence: 1 atggttaacaaagaattttt.....ccgctatcgtcgaatattaa 159

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_estl1:*
11: gb_estl2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	36.8	23.1	398	10	AA812069
C 2	36.8	23.1	399	10	AA426563
C 3	36.8	23.1	628	11	BC938678
C 4	36.8	23.1	974	11	BF794164
5	35.6	22.4	704	11	BG169535
6	35.2	22.1	290	11	M86125
7	35.2	22.1	564	13	AQ372904
8	34.2	21.5	277	10	AA684791
C 9	33.4	21.0	328	10	AW903175
C 10	33.4	21.0	346	11	BE812146
11	33.4	21.0	510	10	A1940791
12	33.4	21.0	572	10	A1940804

13	33.4	21.0	659	10	AU137533
C 14	33.2	20.9	856	13	CNS02PPI
15	32.8	20.6	633	13	BH081489
16	32.6	20.5	518	13	AQ251087
17	32.6	20.5	707	13	AZ369151
C 18	32.2	20.3	385	13	AF295591
C 19	32	20.1	225	13	BH6173
C 20	32	20.1	905	13	BH136332
C 21	31.8	20.0	525	13	B65550
C 22	31.8	20.0	721	13	AZ268413
23	31.8	20.0	796	13	BH115293
24	31.8	20.0	910	13	BH136096
C 25	31.6	19.9	849	10	AA896531
C 26	31.6	19.9	859	13	BH136628
C 27	31.4	19.7	525	13	AQ921505
C 28	31.4	19.7	599	13	FR0031459
C 29	31.2	19.6	777	13	CNS0738Y
C 30	31.2	19.6	1102	13	CNS040LM
C 31	31	19.5	625	10	AV745862
C 32	30.8	19.4	465	10	AW515085
C 33	30.8	19.4	662	13	AQ924679
C 34	30.6	19.2	383	10	AL373357
35	30.6	19.2	517	10	AW444057
C 36	30.6	19.2	1101	13	CNS00GIC
C 37	30.4	19.1	246	10	AV273947
C 38	30.4	19.1	755	13	AQ916515
C 39	30.4	19.1	880	13	CNS013QG
C 40	30.2	19.0	328	10	AI882653
C 41	30.2	19.0	347	10	AA054203
42	30.2	19.0	366	13	AQ247318
43	30.2	19.0	386	11	D85814
44	30.2	19.0	398	13	AZ640930
45	30.2	19.0	508	10	BE239336

ALIGNMENTS

RESULT 1
AA812069/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES

AA812069 398 bp mRNA EST 19-FEB-1998
ob75b07.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1337173 3'
similar to contains element MER22 repetitive element ; , mRNA
sequence.
AA812069
AA812069.1 GI:2881680
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 398)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 2050 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 397.
Location/Qualifiers
1. .398
/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone="IMAGE:I337173"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pW7T3D-Pac (Pharmacia) with a modified
polylinker: Site_1: Not I; Site_2: Eco RI; 1st strand
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD+),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(GT) primer
[5'-TGTTGCAATCTGAAGTGGAGCGCGCCCTCATTTTTTTTTTTT-3',
1. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pW7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
136 a 75 c 61 g 125 t 1 others

```

	Query Match	23.1%	Score 36.8	DB 10	Length 398
	Best Local Similarity	58.0%	Prod No. 0.88	Mismatches 47	Indels 0
	Matches 65	Conservative 0			Gaps 0
QY	15	atttttaaaactaaacttgagtttcagatatgtacgctcagaaactcatagatgaggc	74		
Db	347	ATTAAATAATGTGTATGTGTACATATATACACACATACACATATATTATGC	288		
QY	75	acaggcgcatgaaataggttgtacgacctatttatccaaaaaacttcgagaa	126		
Db	287	ATGTGGTCAAAAGAAATTTGGCTACATAGGGGATTTTCTGACACATCGAAAAA	236		

RESULT	2	
LOCUS	AA426563.3/C	
DEFINITION	399 bp mRNA	
ACCESSION	AA426563	EST 16-OCT-1997
VERSION	2w0290t.s1	Soares-NhHMPu_s1 Homo sapiens CDNA clone IMAGE:768150
KEYWORDS	AA426563	3', mRNA sequence.
SOURCE	AA426563.1	GI:2106818
ORGANISM	human.	
REFERENCE	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
	1 (bases 1 to 399).	

TITLE	COMMENT
WashU-Merck EST Project 1997 Unpublished (1997)	<p>Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu</p> <p>This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -41m13 fwd. ET from Amersham High quality sequence stop: 332.</p>

```

1. .399
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:768150"
/clone_lib="Soares_NhmPU_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
/lab_host="DH10B"

```

```

/Note="Organ: (see below); Vector: p7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPu, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
136 a 77 C 65 g 121 t

BASE COUNT      136 a      77 C      65 g      121 t
ORIGIN

Query Match      23.1%; Score 36.8; DB 10; Length 399;
Best Local Similarity 58.0%; Pred. No. 0.88;
Matches 65; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 15 attttaaaaactaaacttgagtgctcagatatgtcagctcagaaactcatagatgagcc 74
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 335 ATTAAAAATRAATCTGTATCTGTGTACATATATACACACATACATATATTATTC 74
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 75 acagggcgatgaaatagggtgtacgacctattatccaaaaacttgcagaa 126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 275 ATGTGGTGAAGAAATTTGGGTAGATAGGGGATTTTCTGAACACTGCCAAA 224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
BG938678/c
LOCUS
DEFINITION
ACCESSION
BG938678 628 bp mRNA EST 11-JUN-2001
cn27f10.x1 Normal Human Trabecular Bone Cells Homo sapiens CDNA
clone NHTBC_cn27f10 random, mRNA sequence.

```

RESULT	3	
BG938678/c		
LOCUS		
DEFINITION	628 bp mRNA	11-JUN-2001
ACCESSION	cn27f10.x1 Normal Human Trabecular Bone Cells	
VERSION	clone HNPBC_cn27f10 random, mRNA sequence.	
KEYWORDS	BG938678.1 GI:14338050	
SOURCE	EST.	
ORGANISM	human.	
REFERENCE	Homo sapiens	
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. (bases 1 to 628)	
TITLE	Jia,L.B., Young,M.F., Touchman,J.W., Bouffard,G.G., Beckstrom-Stenberg,S.M., Green,E.D., Powell,J.I., Yang,L.M., Robey ,P.G., Hotchkiss,R.N. and Francomano,C.A. SAG: the Skeletal Genome Anatomy Project	
JOURNAL	Unpublished (1997)	
COMMENT	Contact: Libin Jia Medical Genetics Branch National Human Genome Research Institute 10710C101, 9000 Rockville Pike, Bethesda, MD 20892-1267, USA Tel: 301-402-4877 Fax: 301-496-7157 Email: libin@helix.nih.gov	
FEATURES	DNA Sequencing and analyses by National Institutes of Health Intramural Sequencing Center (NISC). Plate: 27 row: f column: 10 Seq primer: -21MI3 forward primer (ABI).	

[illegible]

```

1. 1028
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NHTBC_Cn27f10"
/clone_lib="Normal Human Trabecular Bone Cells"
/sex="Female"
/tissue_type="Bone"
/cell_type="Trabecular Bone Cells"
/lab_host="SURE"
/notes="Organ: Hip; Vector: pBluescript; site_1: EcoRI"
183 a 149 c 117 g 179 t

```

Query Match

```

mRNA sequence.
ACCESSION      BG169535
VERSION        BG169535.1
KEYWORDS       GI:12676238
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
               1 (bases 1 to 704)
               NIH-MGC http://mgc.nci.nih.gov/.
               National Institutes of Health, Mammalian Gene Collection (MGC)
               Unpublished (1999)
               Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-re@mail.nih.gov
               Tissue Procurement: ATCC
               cDNA Library Preparation: Life Technologies, Inc.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLAM10168 row: n column: 14
               High quality sequence stop: 702.
FEATURES       Location/Qualifiers
               1..704
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone_xref="IMAGE:4424653"
               /clone_lib="NIH_MGC_89"
               /tissue_type="hypertrophied, cell line"
               /lab_host="DH10B (phage-resistant)"
               /note="Organ: kidney; Vector: pCMV-Sport6; Site_1: NotI;
               Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
               Average insert size 1.3 kb. Library enriched for
               full-length clones and constructed by Life Technologies.
               Note: this is a NIH_MGC Library."
BASE COUNT    218 a 115 c 151 g 220 t
ORIGIN
: Query Match 22.4%; Score 35.6; DB 11; Length 704;
: Best Local Similarity 58.5%; Pred. No. 2.1;
: Matches 62; Conservative 0; Mismatches 44; Indels 0; Gaps
Qy 21 aaaaactaaacttgagtggttcagatatgcgtccagaaactcatagatgagcacagg 80
   || || || || || || || || || || || || || || || || || || || ||
Db 3 AATAAATGCTGATGCTGTACATATATATACACACACATACATATATATATCATCTGG 62
Qy 81 cgatgaataatgagtggttcacgacctattatccaaaacttcgcagaa 126
   || || || || || || || || || || || || || || || || || || || ||
Db 63 TGAAGAAGATTTGGCTAGTAGAGGGGATTTTCTGAACACATGCACAAA 108
RESULT 6
LOCUS          M86125
DEFINITION     EST072651 Fetal brain, Stragene (cat#936206) Homo sapiens cDNA
               Clone HFBDJ72, mRNA sequence.
ACCESSION      M86125
VERSION        M86125.1
KEYWORDS       GI:274776
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
               1 (bases 1 to 290)
               Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M.,
               Utterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C.
               Sequence identification of 2,375 human brain genes
               Nature 355, 632-634 (1992)
               92168112
               Contact: Kerlavage, AR
               Bioinformatics

```

The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org

Seq primer: M13 Forward.
Location/Qualifiers
1..290

FEATURES

source

/organism="Homo sapiens"
/db_xref="ATCC (inhost):81702"
/db_xref="GDB:D05813E"
/db_xref="taxon:9606"
/clone="HFDU72"
/note="Vector: Fetal brain, Stratagene (cat#936206)"
oligo-dt + random primed cDNA synthesis; lambdaZAP-II
vector, 1.0kb average insr size."

BASE COUNT 86 a 34 c 67 g 102 t 1 others
ORIGIN

Query Match 22.1%; Score 35.2; DB 11; Length 290;
Best Local Similarity 57.1%; Pred. No. 2.5;
Matches 64; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
Qy 15 atttttaaaactaaacttgattcagatgtacgtcagaactcatagatgagc 74
Db 156 ATTTAAATAATGTTGTGTGTATATATATATACACACATACATATATTATGC 215
Qy 75 acagggcgatgaatggttgatcagcctatttccaaactcagaa 126
Db 216 ATGTGGTGAAGAAGATTGGCTAGATAGGGGATTTCTTGACACTGCAAAA 267

RESULT

7

LOCUS

DEFINITION

Q372904

Q372904

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

/clone="RPCI-11-158H17"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"
BASE COUNT 169 a 110 c 147 g 138 t
ORIGIN

Query Match 22.1%; Score 35.2; DB 13; Length 564;
Best Local Similarity 58.7%; Pred. No. 2.7;
Matches 61; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
Qy 14 aatttttaaaactaaacttgattcagatgtacgtcagaactcatagatgagc 73
Db 328 AATTGTGTAACAGAAATGACCTGATCAGATTGTGATTTTAAAGGTATAGACTTGA 387
Qy 74 cacagggcgatgaatggttgatcagcctatttccaaaaa 117
Db 388 ATCAGGCAGACCAATTAACTGAGATGCCACAAATTATCCAAACA 431

RESULT

8

LOCUS

DEFINITION

Q372904

Q372904

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org

Seq primer: M13 Forward.
Location/Qualifiers
1..290

/organism="Homo sapiens"
/db_xref="ATCC (inhost):81702"
/db_xref="GDB:D05813E"
/db_xref="taxon:9606"
/clone="HFDU72"
/note="Vector: Fetal brain, Stratagene (cat#936206)"
oligo-dt + random primed cDNA synthesis; lambdaZAP-II
vector, 1.0kb average insr size."

BASE COUNT 86 a 34 c 67 g 102 t 1 others
ORIGIN

Query Match 22.1%; Score 35.2; DB 11; Length 290;
Best Local Similarity 57.1%; Pred. No. 2.5;
Matches 64; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
Qy 15 atttttaaaactaaacttgattcagatgtacgtcagaactcatagatgagc 74
Db 156 ATTTAAATAATGTTGTGTGTATATATATATACACACATACATATATTATGC 215
Qy 75 acagggcgatgaatggttgatcagcctatttccaaactcagaa 126
Db 216 ATGTGGTGAAGAAGATTGGCTAGATAGGGGATTTCTTGACACTGCAAAA 267

RESULT

7

LOCUS

DEFINITION

Q372904

Q372904

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org

Seq primer: M13 Forward.
Location/Qualifiers
1..290

/organism="Homo sapiens"
/db_xref="ATCC (inhost):81702"
/db_xref="GDB:D05813E"
/db_xref="taxon:9606"
/clone="HFDU72"
/note="Vector: Fetal brain, Stratagene (cat#936206)"
oligo-dt + random primed cDNA synthesis; lambdaZAP-II
vector, 1.0kb average insr size."

BASE COUNT 86 a 34 c 67 g 102 t 1 others
ORIGIN

Query Match 22.1%; Score 35.2; DB 11; Length 290;
Best Local Similarity 57.1%; Pred. No. 2.5;
Matches 64; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
Qy 15 atttttaaaactaaacttgattcagatgtacgtcagaactcatagatgagc 74
Db 156 ATTTAAATAATGTTGTGTGTATATATATATACACACATACATATATTATGC 215
Qy 75 acagggcgatgaatggttgatcagcctatttccaaactcagaa 126
Db 216 ATGTGGTGAAGAAGATTGGCTAGATAGGGGATTTCTTGACACTGCAAAA 267

RESULT

7

LOCUS

DEFINITION

Q372904

Q372904

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

Q372904.1

The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org

Seq primer: M13 Forward.
Location/Qualifiers
1..290

/organism="Homo sapiens"
/db_xref="ATCC (inhost):81702"
/db_xref="GDB:D05813E"
/db_xref="taxon:9606"
/clone="HFDU72"
/note="Vector: Fetal brain, Stratagene (cat#936206)"
oligo-dt + random primed cDNA synthesis; lambdaZAP-II
vector, 1.0kb average insr size."

BASE COUNT 86 a 34 c 67 g 102 t 1 others
ORIGIN

Query Match 22.1%; Score 35.2; DB 11; Length 290;
Best Local Similarity 57.1%; Pred. No. 2.5;
Matches 64; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
Qy 15 atttttaaaactaaacttgattcagatgtacgtcagaactcatagatgagc 74
Db 156 ATTTAAATAATGTTGTGTGTATATATATATACACACATACATATATTATGC 215
Qy 75 acagggcgatgaatggttgatcagcctatttccaaactcagaa 126
Db 216 ATGTGGTGAAGAAGATTGGCTAGATAGGGGATTTCTTGACACTGCAAAA 267

RESULT

7

LOCUS

DEFINITION

RESULT 11

AI940791 510 bp mRNA EST 03-AUG-1999
 LOCUS CM0-ST0050-220799-029-f05 ST0050 Homo sapiens cDNA, mRNA sequence.
 DEFINITION AI940791
 ACCESSION AI940791.1 GI:5687772
 VERSION EST.
 KEYWORDS human.
 SOURCE Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 510)

REFERENCE

HCGP <http://www.ludwig.org.br/ORESTES>
 TITLE The FAPESP/LICR Human Cancer Genome Project
 JOURNAL Unpublished (1999)
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

AUTHORS

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM0&t2=CM0-ST0050-220799-029-f05&t3=1999-07-22&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 85
 High quality sequence stop: 504.

FEATURES

Location/Qualifiers
 1..510
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="ST0050"
 /dev_stage="Adult"
 /note="Organ: Stomach; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

Source

187 a 107 c 132 g 116 t
 Query Match
 Best Local Similarity 21.0%; Score 33.4; DB 10; Length 510;
 Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 57 gaactcatagatgagccacagcgcatgaaataggtgtacgacctattatccaaaa 116
 Db 66 GAGACTCTTTGATGATACACGGGAGCAGAAAGGTGCTGCACTATCTGGCAATCCAGAA 125
 QY 117 acttcagacagctatcacagcc 139
 Db 126 ACCTGCAGACCTTGTCTCGGCACC 148

RESULT 12

AI940804 572 bp mRNA EST 03-AUG-1999
 LOCUS CM0-ST0050-260799-031-f05 ST0050 Homo sapiens cDNA, mRNA sequence.
 DEFINITION AI940804
 ACCESSION AI940804
 VERSION AI940804.1 GI:5687785
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 572)

REFERENCE

HCGP <http://www.ludwig.org.br/ORESTES>
 TITLE The FAPESP/LICR Human Cancer Genome Project
 JOURNAL Unpublished (1999)
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

AUTHORS

HCGP <http://www.ludwig.org.br/ORESTES>
 TITLE The FAPESP/LICR Human Cancer Genome Project
 JOURNAL Unpublished (1999)
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project.

This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM0&t2=CM0-ST0050-260799-031-f05&t3=1999-07-26&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 28
 High quality sequence stop: 571.

FEATURES

Location/Qualifiers
 1..572
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="ST0050"
 /dev_stage="Adult"
 /note="Organ: Stomach; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT

187 a 107 c 152 g 126 t

ORIGIN

Query Match
 Best Local Similarity 21.0%; Score 33.4; DB 10; Length 572;
 Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
 QY 57 gaactcatagatgagccacagcgcatgaaataggtgtacgacctattatccaaaa 116
 Db 131 GAGACTCTTTCGATGATACACGGGAGCAGAAAGGTGCTGCACTATCTGGCAATCCAGAA 190
 QY 117 acttcagacagctatcacagcc 139
 Db 191 ACCTGCAGACCTTGTCTCGGCACC 213

RESULT 13

AI137533 659 bp mRNA EST 24-OCT-2000
 LOCUS AU137533 PLACE1 Homo sapiens cDNA clone PLACE1006685 5', mRNA
 DEFINITION AU137533
 ACCESSION AU137533
 VERSION AU137533.1 GI:10998072
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 659)
 Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
 Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
 Isogai,T.

TITLE

JOURNAL

COMMENT

HRI human cDNA project
 Unpublished (2000)
 Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yano, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3951
 Fax: 81-438-52-3952
 Email: genomics@hri.co.jp

HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

FEATURES

source Location/Qualifiers

1..859
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE1006685"
/clone_lib="PLACE1"
/tissue_type="placenta"
/note="Vector: pME18SFL3"

BASE COUNT 208 a 126 c 171 g 151 t 3 others

ORIGIN

Query Match 21.0%; Score 33.4; DB 10; Length 659;
Best Local Similarity 62.7%; Pred. No. 9.3;
Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 57 gaaactcatagatgagcagcgcatgaaataggtgtgacgacctatttatccaaa 116

DB 284 GACACTCTTCGATCATACAGCGGACAGCAAGAGGTGCTGCACATCTCGCAATCCAGAA 343

QY 117 acttcgagaacgtcatcacgc 139

DB 344 ACCTGCAGACCTTGCTCGGCACC 366

RESULT 14

CNS02PPI/c

LOCUS

DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 156A20 of library G from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

CNS02PPI 856 bp DNA GSS 14-MAY-2000
Tetraodon nigroviridis genome survey sequence T7 end of clone 156A20 of library G from Tetraodon nigroviridis, genomic survey sequence.

REFERENCE

AUTHORS

Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissbach, J.

TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

JOURNAL

REFERENCE

AUTHORS

2. (bases 1 to 856)
Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissbach, J.

TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetraodon>.

FEATURES

source

1..856
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="156A20"
/clone_lib="G"
/note="Genoscope sequence ID : COAG156BA10LPI-end : T7"

BASE COUNT 247 a 134 c 214 g 250 t 11 others
ORIGIN

Query Match

Best Local Similarity 20.9%; Score 33.2; DB 13; Length 856;
Matches 71; Conservative 1; Mismatches 64; Indels 0; Gaps 0;

QY 19 ttaaaactaaacttgagttcagatagtcagctcagaaactcatagatgagggcacg 78

DB 264 TTAATAATATATAGAAATATATATCAATCCATTCATTAGGAATTGATATTATTATACAG 205

QY 79 ggcgatgaaaataggtgtgacgacctatttatccaaaacttcgagaacgtcatcacgc 138

DB 204 GAAAGGAAAGAGGCTGAAACACACACCCCTTTCATAAACTTCTAGTTGGSCCCCCCT 145

QY 139 cccgctatcgtcgaat 154

DB 144 CCTTCTGCCCTAATAT 129

RESULT 15

BH081489

LOCUS

DEFINITION BH081489 633 bp DNA GSS 18-JUL-2001
RPCI-24-321117.TV RPCI-24 Mus musculus genomic clone RPCI-24-321117, DNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Other-GSSs:

Contact:

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html

Plate: 321 row: 1 column: 17

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1..633

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-24-321117"

/clone_lib="RPCI-24"

/sex="Male"

/cell_type="Spleen/Brain"

/note="Vector: pTABAC1; site_1: BamHI; site_2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTABAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT

ORIGIN

190 a 93 c 171 g 179 t

us-09-407-804a-8.rst

Page 8

	Query Match	20.6%	Score 32.8	DB 13	Length 633
	Best Local Similarity	61.9%	Pred. No. 14		
	Matches 52	Conservative 0	Mismatches 32	Indels 0	Gaps 0
QY	56	agaaactcatgatgagggcagggcgatgaaataaggttgacgaacctattatccaaa	115		
Db	257	AGAGACTCATAGGTCAGCTACAGGATTTGTANTAGCTATTAGGATAGATTTTCCAAT	316		
QY	116	aacttgcgaacgctcacacgcgc	139		
Db	317	GACTGTGTTGCCACATTCATCCC	340		

Search completed: March 1, 2002, 14:32:19
Job time: 2623 sec

STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/650.275
FILING DATE: Filed Herewith
ATTORNEY/AGENT INFORMATION:
NAME: Luther, Barbara J.
REGISTRATION NUMBER: 33,954
REFERENCE/DOCKET NUMBER: PF-0067 US
TELEPHONE: 415-855-0555
TELEFAX: 415-852-0195
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 394 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY: LVENNOT01
CLONE: 350290
US-08-650-275-12

Query Match 18.1%; Score 28.8; DB 1; Length 394;
Best Local Similarity 63.9%; Pred. No. 0.63;
Matches 39; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
Qy 72 ggcacagggcgatgaataagggtgtacgacctattatcccaaaacttgcaaacgtca 131
Db 384 GNC A A A N C G N A T G T A A A A A G G T T A T T G G A C C T T T T T C N C C T A T T A T T G C C G A G C T T C N 325
Qy 132 t 132
Db 324 T 324

RESULT 3
US-09-181-318-12/c
Sequence 12, Application US/09181318
Patent No. 6001632
GENERAL INFORMATION:
APPLICANT: Braxton, Scott Michael
APPLICANT: Murry, Lynn E.
TITLE OF INVENTION: HUMAN PROTEIN DISULFIDE ISOMERASE
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/181.318
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/650.275
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Luther, Barbara J.

REGISTRATION NUMBER: 33,954
REFERENCE/DOCKET NUMBER: PF-0067 US
TELEPHONE: 415-855-0555
TELEFAX: 415-852-0195
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 394 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY: LVENNOT01
CLONE: 350290
US-09-181-318-12

Query Match 18.1%; Score 28.8; DB 3; Length 394;
Best Local Similarity 63.9%; Pred. No. 0.63;
Matches 39; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
Qy 72 ggcacagggcgatgaataagggtgtacgacctattatcccaaaacttgcaaacgtca 131
Db 384 GNC A A A N C G N A T G T A A A A A G G T T A T T G G A C C T T T T T C N C C T A T T A T T G C C G A G C T T C N 325
Qy 132 t 132
Db 324 T 324

RESULT 4
US-07-991-867B-21
Sequence 21, Application US/07991867B
Patent No. 5476781
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Hall, Richard L.
APPLICANT: Gruidl, Michael E.
TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/991.867B
FILING DATE: 12-DEC-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 92/14818
FILING DATE: 12-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827.685
FILING DATE: 30-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/657.584
FILING DATE: 19-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UFI14.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800

Db 1917 TTACAAATTAATAAATTTATGTAGATACGACGAAATAAAATTTATATTGGAAGAAA 1976
QY 77 agggcgatgaatagtgtagacatttattccaaacttgcagaacgtctacac 136
Db 1977 TTGAAGCAGAAATATAGATCGGAGACAGATTTCCACGACGCTGTAATAAAC 2036
QY 137 gcc 139
Db 2037 ACC 2039

RESULT 9
US-08-544-332-33
; Sequence 33, Application US/08544332
; Patent No. 5935777
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gerard H. Bencen
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/544,332
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/991,867
; FILING DATE: 07-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/107,755
; FILING DATE: 19-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 92/14818
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,685
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bencen, Gerard H.
; REGISTRATION NUMBER: 35,746
; REFERENCE/DOCKET NUMBER: UFI14.C4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3907 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-544-332-33

Query Match 18.0%; Score 28.6; DB 2; Length 3907;
Best Local Similarity 52.0%; Pred. No. 1.7;
Matches 64; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 17 ttttaaaactaaacttgagtgttcagatatgtacgtctcagaaactcatagatgagcac 76
Db 1917 TTTACAAATTAATAAATTTATGTAGATACGACGAAATAAAATTTATATTGGAAGAAA 1976
QY 77 agggcgatgaatagtgtagacatttattccaaacttgcagaacgtctacac 136
Db 1977 TTGAAGCAGAAATATAGATCGGAGACAGATTTCCACGACGCTGTAATAAAC 2036
QY 137 gcc 139
Db 2037 ACC 2039
RESULT 10
US-08-107-755A-1
; Sequence 1, Application US/08107755A
; Patent No. 5721352
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: U.S.A.
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/107,755A
; FILING DATE: 19-AUG-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,658
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UFI14.C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6768 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Amsacta moorei entomopoxvirus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (65..1459)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1474..2151
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (2239..2475)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2502..2987

```

;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3080...6091
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (6277...6768)
;
US-08-107-755A-1

Query Match      18.0%; Score 28.6; DB 1; Length 6768;
Best Local Similarity 52.0%; Pred. No. 2.1;
Matches 64; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 17 ttttaaaactaaacttgagttgcagatgtacgctcagaactcatagatgaggcac 76
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4191 TTTACAAATTAATAAATTTATGTAGATAACGACGAAATATAATTTATTTGGAAGAA 4250
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 77 agggcgatgaaatagggtgtacgacctattatccaaaacttcgagaacgtcatcac 136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4251 TTGAAGCAGATATAGATGCCGAAGACAAAGTATTCACGAACGTGTAATAAAC 4310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 137 gcc 139
Db 4311 ACC 4313

```

```

RESULT 11
US-07-991-867B-1
; Sequence 1, Application US/07991867B
; Patent No. 5476781
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/991.867B
; FILING DATE: 12-DEC-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 92/14818
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,685
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF114.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-3800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8457 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

```

```

;
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Amsacta moorei entomopoxvirus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (65...1459)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1474...2151
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (2239...2475)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2502...2987
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3080...6091
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (6277...6768)
;
US-07-991-867B-1

```

```

Query Match      18.0%; Score 28.6; DB 1; Length 8457;
Best Local Similarity 52.0%; Pred. No. 2.2;
Matches 64; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 17 ttttaaaactaaacttgagttgcagatgtacgctcagaactcatagatgaggcac 76
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4191 TTTACAAATTAATAAATTTATGTAGATAACGACGAAATATAATTTATTTGGAAGAA 4250
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 77 agggcgatgaaatagggtgtacgacctattatccaaaacttcgagaacgtcatcac 136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4251 TTGAAGCAGATATAGATGCCGAAGACAAAGTATTCACGAACGTGTAATAAAC 4310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 137 gcc 139
Db 4311 ACC 4313

```

```

RESULT 12
US-08-544-332-1
; Sequence 1, Application US/08544332
; Patent No. 5935777
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gerard H. Bencen
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/544,332
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/991,867
; FILING DATE: 07-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/107,755

```


Query Match 17.7%; Score 28.2; DB 3; Length 5340;
Best Local Similarity 54.3%; Pred. No. 2.6;

Best local similarity 32.0%, P.I.D.NO. 2.2;
Matches 64; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
17 ttttaaaaaactaaacttgagtgttctcagatatgtacgcgtcaaaaaactcatatagatgagccac 76

RESULT 14
US-08-423-752-9
; Sequence 9, Application US/08423752

```

: GENERAL INFORMATION:
:
: APPLICANT: Takashi OKADO et al.
:
: TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING
:
: TITLE OF INVENTION: AUREOBASIDIN SENSITIVITY
:
: NUMBER OF SEQUENCES: 22
:
: CORRESPONDENCE ADDRESS:
:

```

:
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: Wenderoth, Lind & Ponack
:
: STREET: 805 Fifteenth Street, N.W., #700
:
: CITY: Washington
:
: STATE: D.C.
:
: COUNTRY: U.S.A.
:

```

? COUNTRY: U.S.A.
? ZIP: 20005
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: MS-DOS
? SOFTWARE: Wordperfect 5.1
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/423.752
? FILING DATE: April 18, 1995
?
```

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/243,403
; FILING DATE: May 16, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5340
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-423-752-9

```

```

Query Match      17.7%; Score 28.2; DB 3; Length 5340;
Best Local Similarity 54.3%; Pred. No. 2.6;
Matches 57; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 13 gaatttttaaaactaaacttgagtggttcagatatgtacgctcagaaactcatagatgag 72
Db 753 GATATTTCCCTCAGAACGCTGTTTCATCTTGCACCTCTAAGAAATTCAGAAAGTA 812
QY 73 gcacagggcgatgaaataggtgttacgacctattttatccaaaaa 117
Db 813 CCACAAACCGATGACGAGAGGAGATATATCTCTCTGTTCCATACA 857

```

```

RESULT 15
US-08-945-994-2
; Sequence 2, Application US/08945994
; Patent No. 6043051
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: PROMOTER
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/945,994
; FILING DATE: No. 6043051ember 6, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5340

```

```

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-945-994-2

```

```

Query Match      17.7%; Score 28.2; DB 3; Length 5340;
Best Local Similarity 54.3%; Pred. No. 2.6;
Matches 57; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 13 gaatttttaaaactaaacttgagtggttcagatatgtacgctcagaaactcatagatgag 72
Db 753 GATATTTCCCTCAGAACGCTGTTTCATCTTGCACCTCTAAGAAATTCAGAAAGTA 812
QY 73 gcacagggcgatgaaataggtgttacgacctattttatccaaaaa 117
Db 813 CCACAAACCGATGACGAGAGGAGATATATCTCTCTGTTCCATACA 857

```

Search completed: March 1, 2002, 13:49:41
Job time: 65 sec

THIS PAGE BLANK (USPTO)